DOCUMENT RESUME

ED 333 048 TM 016 617

AUTHOR	Linacre, John Michael; And Others
TITLE	Criterion-Sensitive Measurement: A Study in Multi-Dimensionality.
SPONS AGENCY	Centers for Disease Control (DHHS/PHS), Atlanta, GA.; National Inst. on Disability and Rehabilitation Research (ED/OSERS), Washington, DC.
PUB DATE	Apr 91
NOTE	11p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Criterion Referenced Tests; *Disabilities; Evaluation Criteria; *Patients; Physical Disabilities; *Rating Scales; *Rehabilitation
IDENTIFIERS	*Functional Independence Measure (Forer et al); Multidimensional Discrimination; Rasch Model

ABSTRACT

The Functional Independence Measure (FIM) of S. Forer and others (1987) records the degree of disability of rehabilitation patients between "Total Dependence" and "Complete Independence." Using the FIM, ratings on a seven-point scale are made by therapists and other expert care-providers at the time of patient admission to rehabilitation, at the time of patient discharge from rehabilitation, and at a follow-up assessment several weeks after patient discharge. The effectiveness of the rehabilitation program is measured by comparing the FIM scores for a patient at admission and at discharge. Rasch analysis of the admission or discharge FIM ratings of a sample of 33,646 patients shows that the 18 items of the FIM do not define a single statistical variable. Analysis highlights two statistically and substantively different variables: (1) disability in motor functions; and (2) disability in cognitive functions. Separate Rasch analysis of these two variables confirms this conclusion and suggests further refinements that would increase FIM usefulness. Three tables contain study data. The FIM is presented. (Author/SLD)

Criterion-sensitive Measurement

A study in multi-dimensionality

by

John Michael Linacre
and
Benjamin D. Wright

MESA Psychometric Laboratory
Department of Education
University of Chicago

Allen W. Heinemann
Rehabilitation Institute of Chicago

Byron Hamilton
State University of New York at Buffalo

Paper presented at
American Educational Research Association Annual Meeting
Chicago, Illinois
April 1991

Abstract

The Functional Independence Measure (FIM) records the degree of disability of rehabilitation patients between "Total Dependence" and "Complete Independence". Rasch analysis of the admission or discharge FIM ratings of a sample of 33,646 patients shows that the 18 items of the FIM do not define a single statistical variable. Analysis brings out two statistically and substantively different variables: disability in motor functions and disability in cognitive functions. Separate Rasch analysis of these two variables confirms this conclusion, and suggests further refinements which would increase FIM usefulness.

Key words: Rasch Measurement, Unidimensionality, Multidimensionality, Criterion-referenced testing, Functional assessment, Rating scales

Introduction

Since there is usually more than one way to reach a developmental criterion, measuring instruments are then sought which match alternative paths of development. A key question is to what extent measuring instruments can be adapted to the characteristics of the persons assessed without confusing the meaning of the criterion.

the Functional Independence Measure (FIM, Forer et al. 1987), used in rehabilitation medicine for the assessment of disability, provides an example of this. The FIM is intended to mark the patient's location between total dependency on others and completely independent functioning.

The Functional Independence Measure (FIM)

The FIM consists of 18 items, each rated on a seven-point scale (see Figure 1). 3 of the items have alternate forms e.g. Item L is rated for locomotion either by walking or by wheelchair. Ratings on the seven-point scale are made by therapists and other expert care-providers at admission to rehabilitation, at discharge from rehabilitation, and at a follow-up assessment several weeks after discharge.

The effectiveness of the rehabilitation program is measured by comparing the FIM scores for a patient at admission and discharge. Degree of functioning is manifestly multi-dimensional. Compare a youth entering rehabilitation for a gun-shot wound to the spine, with a geriatric stroke victim. Identical FIM scores for these patients would indicate different forms of dependency. Some adjustment must be made for the differences in the

---

This study was supported by grants from the Centers for Disease Control, Center for Environmental Health and Injury control, Division of Injury Epidemiology and Control (Howard Hill, Program Officer), and from the U.S. Department of Education, National Institute on Disability and Rehabilitation Research.

nature of the impairment of these two patients. If, however, an extreme position is adopted in which each patient defines a unique path of rehabilitation, then there is no longer any basis on which to compare patient progress, therapy effectiveness, or on which to allocate scarce resources.

The measurement challenge in this instance is to construct a meaningful and useful partition of the multi-dimensional space implied by the application of the 18 item FIM to the 13 impairment (ICG) groups listed in Table 1. Each partition of the space must function as a clearly defined unidimensional variable. But, for generality and utility, the number of unidimensional variables must few and the variables they define must relate directly to clinical practice.

Can the FIM define one variable of "disablement"?

Since the FIM is used to obtain one "basic indicator of severity of disability" (ibid.), the designers intended it to be unidimensional. Dimensionality is always a matter of definition (Stahl 1991), but Rasch analysis (Rasch 1960, 1980) provides a means to determine whether a designer's intention has been fulfilled in a useful way. This is done by fitting FIM observations to a unidimensional measurement model. Misfit of the data to this model locates areas in which empirical unidimensionality has not been achieved and shows how the FIM can be improved.

The Rasch measurement model used in this analysis (Andrich 1978, Wright & Masters 1982) is:

$$\log\left(\frac{P_{nij}}{P_{nij-1}}\right) = B_n - D_i - F_j \qquad j=2,7 \qquad (1)$$

where P_{nij} is the probability of success of patient n, with ability measure B_n, receiving a rating of j on item i (one of the 18 FIM items), with difficulty calibration D_i.
P_{nij-1} is the probability of receiving a rating of j-1 under the same circumstances.
F_j is the extra ability indicated by a rating of category j, relative to that indicated by a rating of category j-1, on the 7 category FIM rating scale.

Admit and discharge FIM ratings for 33,646 patients were obtained from the Uniform Data System for Medical Rehabilitation (UDS) at the State University of New York at Buffalo. These records were used to evaluate the unidimensionality of the FIM. Patient records were assigned alternately to one of two groups. The admission scores for one group and the discharge scores for the other group were Rasch analyzed. The item statistics are shown in Table 2 (BIGSTEPS, Wright & Linacre 1991).

The samples in Table 2 are after the elimination of records with extreme scores (total dependency or complete dependence for all functions) for whom within-patient variance cannot be observed. The fit statistics in Table 2 are the observed-to-modelled mean-square-variance ratios. Values greater than 1 indicate unexplained variance. The outfit statistic sums squared standardized residuals and is more sensitive to individual unexpected outliers. The infit statistic sums information-weighted squared residuals, and is more sensitive to unexpected patterns.

The measurement frame of reference in Table 2 is established by the modal performance of the items. When two heterogeneous groups of items are combined in one test, it is the larger of the two groups which dominates the frame of reference. Items in the other group tend to misfit. Table 2 shows a preponderance of misfit among the later mostly cognitive items.

Certain items display considerable misfit for substantive reasons. Item M, Stairs, has substantial misfit at admission and discharge. This item reports the patient's function going up and down 12 to 14 stairs. When this item is not observable, or when the patient is carried, the FIM instructions are to score the item "1", total assist. Because of the safety considerations associated with this item, clinicians report that they seldom attempt to assess a patient's performance on this item, unless satisfied the patient has a good level of functioning. Consequently this item is used idiosyncratically and cannot be expected to cooperate in defining an overall frame of reference. Nevertheless, its inclusion in the FIM may be useful for patients approaching completely independent functioning. Revised scoring instructions for this item would improve the measurement properties of the FIM.

The pattern of misfit in Table 2 reveals a clear disjunction between the "Motor" and "Cognitive" items. Every cognitive item provokes misfit at both time points. It is clear that this mixture of motor and cognitive items is producing a measuring system in which a particular measure is ambiguous with respect to the definition of the variable. A single characteristic pattern of disablement corresponding to a particular measure on the FIM is not identifiable.

The "Motor" and "Cognitive" variables

In order to obtain statistically valid and clinically meaningful measures from the FIM, we must redefine the FIM in terms of two substantively defined unidimensional variables. Further Rasch analysis is performed following this new definition of the FIM.

Table 3 shows that the measurement capability of the FIM is considerably improved. The cognitive items now show acceptable fit statistics at admission and discharge. A further indication of the improved discrimination of this part of the instrument is the increased range of the cognitive item calibrations. In the combined analysis their range is .47 logits, but in the separate analyses the range becomes twice as large, .95, indicating twice the facility to usefully distinguish between different levels of cognitive disability.

The range of motor item calibrations has also increased from 2.44 to 3.08 logits, a consequence of the removal of the noise introduced into the frame of reference defined by the motor items when the cognitive items are calibrated with them. The misfit of items A (Feeding), G (Bladder), H (Bowel) and K (Tub transfer) has also been brought into focus by the removal of the obscuring cloud of noise produced by including the cognitive items in the original analysis. This further insight into the nature of the FIM shows that further refinement could prove useful.

Further refinement of the FIM variables

The statistically motivated, but substantively chosen, separation of the FIM items into motor and cognitive variables has increased its measurement capability. But there is still some lack of clarity in the interpretation of FIM motor measures.

There are several avenues of attack on this problem. Several FIM items are ambiguous. In addition to the double definition of a rating of "1" as both "Total Assist" and "Not Testable", several items also have double definitions. Item L, "Walk/Wheel Chair", Item N, "Comprehension Auditory/Visual" and Item O "Expression Verbal/Non-verbal" are rated according to the most usual behavior exhibited by the patient. But the alternative forms of these items do not represent equal disability challenges. Combining the two item forms into one item dilutes their measurement capacity. This suggests further analysis in which each item form is treated as a separate item.

The patient population heterogeneous, i.e. patients are disabled in the same way. In fact, differing forms of impairment lead to different patterns of rating on the FIM items. This information is useful, diagnostically, because it helps identify the therapy to be administered to the patient, but, it confounds the measuring process by restricting the generality of a measure. It might be impractical to have a separate measuring system for each of the 13 ICG codes, but a clearly defined, diagnostically convenient, and statistically superior grouping of ICG codes by their measurement characteristics would increase FIM usefulness.

Conclusion

The FIM was intended to quantify a unidimensional disability variable. Analysis of FIM ratings confirms that the FIM has measurement capability at rehabilitation admission and discharge. Analysis also points out that the FIM items detect two substantively different variables of disability, the motor and the cognitive. Separate analysis of the items representing these two variables confirms the hypothesis that two measures provide more meaningful information from the FIM than one. The residual misfit of the data from these two newly defined unidimensional frames of reference suggest that even more meaning can be obtained from FIM ratings.

References

Andrich D (1978) A rating formulation for ordered response categories. Psychometrika 43(4): 69-81.

Forer S, Grainger C, et al. (1987) Functional Independence Measure. Buffalo NY: The Buffalo General Hospital, State University of New York at Buffalo.

Rasch G (1960, 1980) Some probabilistic models for intelligence and attainment tests. Chicago: University of Chicago Press

Stahl J A (1991) Lost in the dimensions. Rasch Measurement 4(4): 120.

Wright B D, Linacre J M (1991) BIGSTEPS Rasch analysis computer program. Chicago: MESA Press

Wright B D, Masters G N (1982) Rating Scale Analysis. Chicago: MESA Press.

FUNCTIONAL INDEPENDENCE MEASURE

FIM

LEVELS		
	7 Complete Independence (Timely, Safely) 6 Modified Independence (Device)	NO HELPER
	Modified Dependence 5 Supervision 4 Minimal Assist (Subject = 75%+) 3 Moderate Assist (Subject = 50%+) Complete Dependence 2 Maximal Assist (Subject = 25%+) 1 Total Assist (Subject = 0%+)	HELPER

	ADMIT	DISCHG	FOL-UP
Self Care			
A. Feeding			
B. Grooming			
C. Bathing			
D. Dressing-Upper Body			
E. Dressing-Lower Body			
F. Toileting			
Sphincter Control			
G. Bladder Management			
H. Bowel Management			
Mobility			
Transfer:			
I. Bed, Chair, W/Chair			
J. Toilet			
K. Tub, Shower			
Locomotion			
L. Walk/wheelChair	wc	wc	wc
M. Stairs			
Communication			
N. Comprehension	a v v n	a v v n	a v v n
O. Expression			
Social Cognition			
P. Social Interaction			
Q. Problem Solving			
R. Memory			
Total			

Note: If item is not testable, enter level 1.

Figure 1. The Functional Independence Measure (FIM)

Impairments	
1	Stroke
2	Brain dysfunction
3	Neurologic conditions - other
4	Spinal cord dysfunction
5	Amputee
6	Arthritis
7	Pain syndromes
8	Orthopaedic conditions
9	Cardiac
10	Pulmonary
11	Burns
12	Congenital deformities
13	Other disabling impairments

Table 1. The Impairments (ICG).

	AT ADMISSION (16545 RECORDS)				AT DISCHARGE (16483 RECORDS)			
FIM ITEM	CALIBRTN	ERROR	INFIT	OUTFIT	CALIBRTN	ERROR	INFIT	OUTFIT
MOTOR ITEMS								
A Feeding	-.74	.01	1.02	.98	-.88	.01	1.06	.87
B Grooming	-.40	.01	.77	.77	-.50	.01	.79	.73
C Bathing	.37	.01	.62	.65	.28	.01	.62	.67
D Dress-Upper	-.15	.01	.76	.75	-.27	.01	.78	.73
E Dress-Lower	.44	.01	.63	.65	.26	.01	.64	.65
F Toileting	.28	.01	.73	.71	.18	.01	.71	.67
G Bladder	-.22	.01	1.38	1.28	-.21	.01	1.46	1.24
H Bowel	-.28	.01	1.18	1.12	-.25	.01	1.20	1.05
I Bed,Chair	.27	.01	.54	.70	.12	.01	.54	.57
J Toilet Tr	.38	.01	.60	.69	.23	.01	.53	.55
K Tub Trfer	.92	.01	1.17	1.19	.77	.01	1.15	1.15
L Walk/Wchr	.49	.01	.99	1.10	.33	.01	.88	.93
M Stairs	1.70	.01	1.87	1.67	1.26	.01	1.57	1.55
COGNITIVE ITEMS								
N Comprehen	-.83	.01	1.35	1.42	-.48	.01	1.63	1.62
O Expressin	-.81	.01	1.45	1.47	-.45	.01	1.81	1.79
P Social In	-.61	.01	1.13	1.20	-.40	.01	1.29	1.24
Q Problem S	-.36	.01	1.17	1.25	.05	.01	1.30	1.29
R Memory	-.45	.01	1.26	1.34	-.05	.01	1.43	1.41

Table 2. Item Difficulties for all 18 FIM items calibrated together. The analysis is of FIM ratings for two samples of patients, one at admission, the other at discharge. Item calibrations (CALIBRTN) are in logits. Infit and outfit statistics are mean-squares with expectation of 1. Values larger than 1 indicate unmodelled variance.

FIM ITEM	AT ADMISSION (16191 RECORDS) CALIBRTN	ERROR	INFIT	OUTFIT	AT DISCHARGE (13009 RECORDS) CALIBRTN	ERROR	INFIT	OUTFIT
MOTOR ITEMS								
A Feeding	-1.25	.01	1.25	1.29	-1.32	.01	1.35	1.24
B Grooming	-.81	.01	.93	.96	-.81	.01	1.01	.94
C Bathing	.18	.01	.68	.70	.24	.01	.72	.80
D Dress-Upp	-.49	.01	.87	.85	-.51	.01	.92	.86
E Dress-Low	.27	.01	.61	.61	.22	.01	.68	.68
F Toileting	.06	.01	.76	.73	.10	.01	.77	.72
G Bladder	-.58	.01	1.67	1.58	-.43	.01	1.84	1.54
H Bowel	-.67	.01	1.40	1.38	-.47	.01	1.49	1.33
I Bed,Chair	.05	.01	.51	.66	.01	.01	.56	.61
J Toilet Tr	.19	.01	.57	.62	.18	.01	.52	.54
K Tub Trfer	.88	.01	1.26	1.21	.91	.01	1.32	1.31
L Walk/Wchr	.33	.01	1.14	1.28	.31	.01	1.10	1.20
M Stairs	1.83	.01	1.90	1.56	1.57	.01	1.68	1.61

FIM ITEM	AT ADMISSION (16065 RECORDS) CALIBRTN	ERROR	INFIT	OUTFIT	AT DISCHARGE (11975 RECORDS) CALIBRTN	ERROR	INFIT	OUTFIT
COGNITIVE ITEMS								
N Comprehen	-.44	.01	1.03	1.02	-.37	.01	1.09	1.07
O Expressin	-.39	.01	1.20	1.07	-.33	.01	1.29	1.13
P Social In	.00	.01	1.22	1.26	-.23	.01	1.17	1.14
Q Problem S	.51	.01	.78	.76	.56	.01	.75	.72
R Memory	.32	.01	.86	.80	.37	.01	.86	.80

Table 3. Item Difficulties of 5 Cognitive and 13 Motor items calibrated separately for the two samples of Table 2.